



**EXAMINING THE STATISTICAL RIGOR OF TEST AND EVALUATION
RESULTS IN THE LIVE, VIRTUAL AND CONSTRUCTIVE ENVIRONMENT**

GRADUATE RESEARCH PROJECT

James G. Wilson, Major, USAF

AFIT/IOA/ENS/11-06

**DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY**

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

AFIT/IOA/ENS/11-06

**EXAMINING THE STATISTICAL RIGOR OF TEST AND EVALUATION RESULTS
IN THE LIVE, VIRTUAL AND CONSTRUCTIVE ENVIRONMENT**

GRADUATE RESEARCH PROJECT

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the
Degree of Master of Science in Operational Sciences

James G. Wilson

Major, USAF

June 2011

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

AFIT/IOA/ENS/11-06

**EXAMINING THE STATISTICAL RIGOR OF TEST AND EVALUATION
RESULTS IN THE LIVE, VIRTUAL AND CONSTRUCTIVE ENVIRONMENT**

GRADUATE RESEARCH PROJECT

James G. Wilson

Major, USAF

Approved:

//SIGNED//
Dr. Raymond R. Hill, Civ, USAF (Advisor)

26 May 11
Date

Abstract

The Department of Defense has mandated that weapons systems undergo persistent and realistic testing in a joint operational environment. Testing for new weapons systems is to occur early and often, in an operationally realistic environment, in order to identify and correct problems before resolution options become technically infeasible and/or cost prohibitive. Executing the appropriate fidelity of testing solely in the live environment is not always a viable course of action. Advances in distributed testing capabilities combined with the establishment of the technical infrastructure are producing a continually expanding group of distributed capable participants able to play a role in robust joint operational scenarios. However, obtaining statistically rigorous results from virtual tests, especially those pertaining to Operational Test and Evaluation, remains elusive. Several considerations associated with the Design of Experiments for virtual testing are outlined, and potential methodologies explored, with the aim to ensure rigorous and actionable results are produced when using live, virtual and constructed simulations for test purposes.

Table of Contents

	Page
Abstract	iv
List of Figures	vii
List of Tables	viii
1. Background	1
1.1 DOT&E Initiatives	1
1.2 Options for Program Managers	3
1.3 Role of JMETC	4
1.4 Role of LVC in Realistic Joint Operational Testing	7
1.5 Way Ahead	9
2. Case Study - AGILE Fire	10
2.1 AGILE Fire - Background	10
2.2 AGILE Fire Phase-I: January 2010	13
2.2 AGILE Fire Phase-II: August 2010	17
2.3 AGILE Fire PHASE III - January 2011	19
2.4 AGILE Fire Summary	21
3. Analysis	22
3.1 Big Picture - Positive Trends	22
3.2 LVC in OT&E	22
3.3 Conflicting Forces: Robust Joint Environment vs. Manageable Design	23
3.4 Developing Experimental Measures	24
3.5 Factor Definition	28

3.6	<i>Test Discipline and Data Collection</i>	32
3.7	<i>Possible DOE Techniques</i>	34
3.7.1	<i>Full Factorial Design</i>	34
3.7.2	<i>Fractional Factorial Designs</i>	34
3.7.3	<i>Orthogonal Arrays</i>	36
3.7.4	<i>Split-Plot Design</i>	37
3.8	<i>Summary</i>	40
Appendix A: Air University Blue Dart		42

List of Figures

	Page
Figure 1 - JMETC Program (JMETC 2011b).....	5
Figure 2 - JMETC Connectivity (JMETC 2011a)	6
Figure 3 - JTEM Methodology (Bjorkman 2008).....	8
Figure 4 - AGILE Fire Battlespace (Feldman 2010)	11
Figure 5 - AGILE Fire Phase-II Battlespace (March 2010)	12
Figure 6 - Airspace Structure (JAGIC 2009).....	14
Figure 7 - Potential JAGIC Configuration (JAGIC 2009).....	15
Figure 8 - Agile Fire Phase-II Close Air Support (Allison 2010).....	18
Figure 9 - NEW Small Diameter Bomb Employment (Watson 2010)	19
Figure 10 - NEW Data Flow (Watson 2010)	20

List of Tables

	Page
Table 1 - AGILE Fire Phase-I Initiatives (March 2009).....	13
Table 2 - NEW Analysis Traceability Matrix (McRae and Watson 2010).....	25
Table 3 - Possible NEW Factors of Interest (NEW Run Matrix 2011)	30
Table 4 - Notional $4^2 \times 3^2 \times 2^1$ Orthogonal Array (Bolboaca and Jantschi 2007)	36
Table 5 - Split-Plot Experiment	39
Table 6 - OT&E LVC Recommendations	40

1. Background

1.1 DOT&E Initiatives

The role of Operational Test and Evaluation is to provide analytically sound information regarding the survivability, effectiveness and suitability of new systems in combat operations (Gilmore 2009). Analytically sound and operationally relevant OT&E results provide civilian and military leadership the basis to make informed procurement decisions about the system of interest. The impact of OT&E processes are especially relevant in the current geopolitical environment with United States military personnel, engaged in combat operations on multiple fronts, relying on Department of Defense (DoD) procured systems to provide them the necessary tools to achieve mission success.

The Department of Operational Test and Evaluation (DOT&E) has issued a directive outlining several initiatives to improve testing across the DoD. An emphasis is now placed on recognizing, through early testing, effective and suitable weapons systems capable of bringing a new capability to the battlefield. These programs will be assessed by DOT&E for accelerated testing and early fielding (Gilmore 2009).

Identifying system performance shortfalls early is also critical if new and reliable systems are to be delivered on-time and within budget. To achieve these ends, DOT&E initiatives contain a persistent theme of conducting testing of new systems in a realistic joint environment. This includes testing system subcomponents under anticipated operational loads and conditions prior to integration into the "full-up" system (Gilmore 2009).

The Defense Science Board and National Academies identified the separation of developmental and operational testing as a notable problem area (Defense Science Board 2008). It is not surprising that a failure to test in an operationally relevant environment until late in development process can mask some critical performance deficiencies. Failing to conduct realistic testing early identifies system issues at a developmental stage where corrections, even if technically possible, often come at an exorbitant cost coupled with delays in the program timeline (Gilmore 2009).

DOT&E is emphasizing Design of Experiments (DOE) during testing to "increase the use of scientific and statistical methods in developing rigorous, defensible test plans and in evaluating their results" (Gilmore 2010). Effective DOE requires the identification, preferably early in development phases, of core experimental factors that when varied produce responses that demonstrate a proposed weapons system will provide an improved military capability. Furthermore, subsequent test results are of limited actionable value unless they are accompanied by an overview outlining the scope of the assessed operational performance envelope accompanied with rigorously established confidence levels (Gilmore 2009).

In order to achieve these ends, DOT&E outlined specific expectations for Test and Evaluation Master Plans (TEMP) (Gilmore 2010). Test plans should include:

1. The goal of the experiment.
2. Quantitative mission oriented response variables for effectiveness and suitability.
3. Identification of factors that measure effectiveness and suitability.
4. Span the pertinent levels of the factors.

5. Methods to strategically vary factors with respect to responses of interest.

6. Rigorous statistical measures of merit (power and confidence).

A final DOT&E initiative is to substantially improve suitability estimates prior to a system entering IOT&E (Gilmore 2009). Statistically rigorous results from carefully designed and operationally realistic tests should produce insights into projected system suitability. Early identification of potential suitability issues prior to IOT&E will provide program managers and appropriate leadership with insights into the likelihood of IOT&E outcomes.

1.2 Options for Program Managers

The requirement for realistic joint testing leaves program managers with three fundamental options (Ferguson and Brown 2010). An ideal option of executing testing in a purely live venue throughout the entire development process is often infeasible. Creating a sufficiently robust joint environment comprised of the necessary C4ISR and tactical assets is, with the current operations tempo and airframe limitations, simply not possible on a persistent basis. Ironically, occasional opportunities such as the Joint Expeditionary Force Experiment exist but often incorporate pre-event testing in the virtual arena as a risk reduction measure (USAF GCIC 2009). A program relying solely on live joint operational test events will likely lack sufficient opportunities to identify system problems throughout the developmental process.

Conversely, conducting virtual testing alone is likely to mask performance and reliability deficiencies that would otherwise be identified during live testing. It is

difficult to envision the approval of a TEMP that lacks live testing. A system ultimately gains credibility with the operator once proven in live conditions, interacting as a live system within systems, under conditions similar to those expected in combat (Ferguson and Brown 2010).

The obvious option is a combination of both virtual and live testing thoughtfully placed throughout the development process. Conducting robust testing in the virtual arena helps program managers execute focused testing early in the design process without waiting for, or integrating into, an infrequent live exercise. This provides program managers the venue to identify and resolve issues at a point in the process where modifications are likely to be technically possible and comparatively cost attractive. Virtual testing can then be re-accomplished on a persistent basis until the point in the developmental process where live testing is necessary (Ferguson and Brown 2010).

1.3 Role of JMETC

The 2004 DoD *Strategic Planning Guidance for Joint Testing in Force Transformation* outlined the need for adequate and realistic joint operational testing and evaluation (T&E) in the development process. This guidance recommended the development of new testing capabilities that evaluate the effectiveness of a new system as part of a capability-based process (Ferguson and Brown 2010). The resulting Testing in a Joint Environment Roadmap recognized that by emphasizing realistic joint testing the strategic guidance becomes unattainable if testing is predominantly reliant on live assets. The roadmap therefore recommended that "a persistent, robust modern networking infrastructure for systems-of-systems engineering, Developmental T&E (DT&E), and

Operational T&E (OT&E) must be developed that connects distributed LVC resources, enables real time data sharing and archiving, and augments realistic OT&E/Initial OT&E of joint systems and systems of systems" (DoD 2004b).

In December, 2005, the DoD directed the establishment of the Joint Mission Environment Test Capability (JMETC). Their role was to create a distributed technical infrastructure with enough flexibility to cost-effectively integrate and configure LVC resources to achieve specific joint operational test requirements (see Figure-1). The resulting Test and Training Enabling Network Architecture (TENA) successfully debuted as a key element in the 2007 INTEGRAL FIRE joint operational test event (Ferguson and Brown 2010).

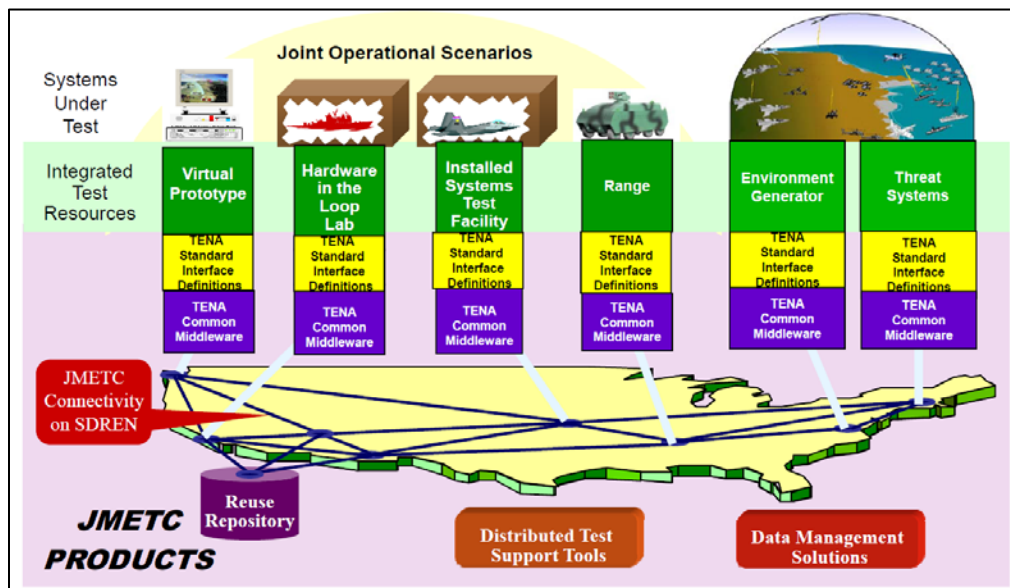


Figure 1 - JMETC Program (JMETC 2011b)

Enhancements to the JMETC capabilities have since resulted in a greater capability of persistently linking joint platforms for test events. This includes the ability to quickly reconfigure network infrastructure to expedite integration of LVC participants,

incorporate systems through connectivity with other DoD networks, and options for on-site customer support expertise (Ferguson and Brown 2010).

The JMETC program is taking several steps to mitigate challenges inherent to LVC testing. In addition to expertise for technical issues, JMETC provides assistance in incorporating distributed test requirements into test planning documents (e.g., TEMP). They also provide expertise in developing test support and data analysis tools, data logging and network performance analysis before, during and after the test. Additionally, a JMETC reuse repository is in place to provide users access to lessons learned and forums providing opportunities for collaboration (Ferguson and Brown 2010).

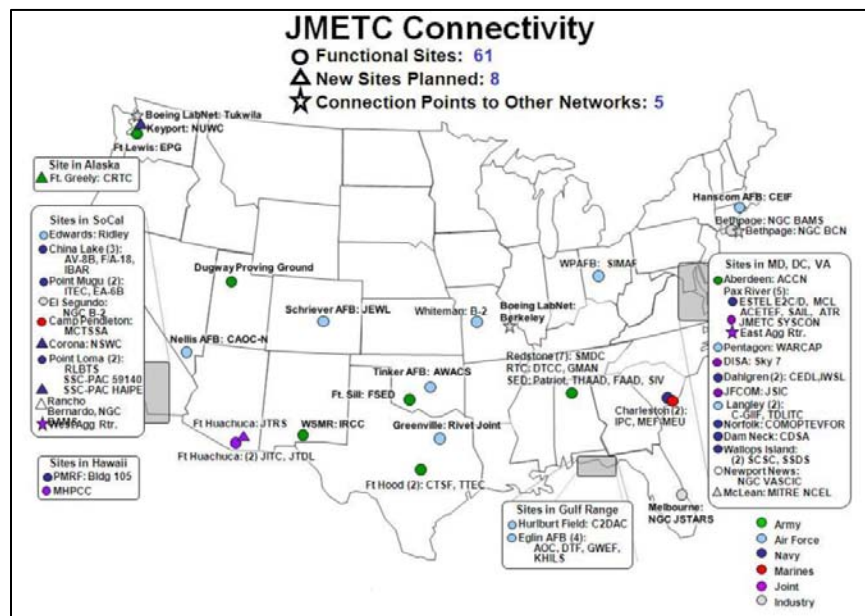


Figure 2 - JMETC Connectivity (JMETC 2011a)

JMETC has activated over 60 sites around the country with the integration of several additional sites planned in 2011 (see Figure-2). This ability to connect systems

that increasingly span the joint operational environment has set the stage for persistent joint T&E in the LVC domain.

1.4 Role of LVC in Realistic Joint Operational Testing

DOT&E established the Joint Test and Evaluation Methodology (JTEM) project in February 2005 with the mandate to investigate, evaluate, and make recommendations to improve the ability to test across the acquisition life cycle in realistic joint mission environments. This guidance included a particular focus on using the LVC joint test environment to evaluate system performance and joint mission effectiveness (Bjorkman and Gray 2009a).

The LVC battlespace consists of simulation resources linked together from geographically separated locations. Constructive elements are computer-generated entities that help fulfill mission scenario requirements (e.g., computer programmed or operator 'driving' simulated F-15E at a computer console). Virtual entities consist of a participant in a simulator trained to employ that specific tactical asset (e.g., F-15E pilot in a simulator) and live entities are manned tactical assets (e.g., airborne F-15E). All participating LVC elements are linked together through the distributed network to form a single operational environment in the virtual battlespace.

JTEM utilized the 2007 INTEGRAL FIRE test, sponsored by the Air Force Integrated Collaborative Environment program, to evaluate their methodology for designing and executing a LVC test event. The INTEGRAL FIRE lessons learned have been incorporated into an updated JTEM capability test methodology and applied in subsequent test events (see Figure-3). The INTEGRAL FIRE Joint Fires virtual

battlespace provided the venue for testing of air-to-surface network-enabled weapons (NEW) and ground-launched surface-to-surface precision attack (Bjorkman and Gray 2009b).

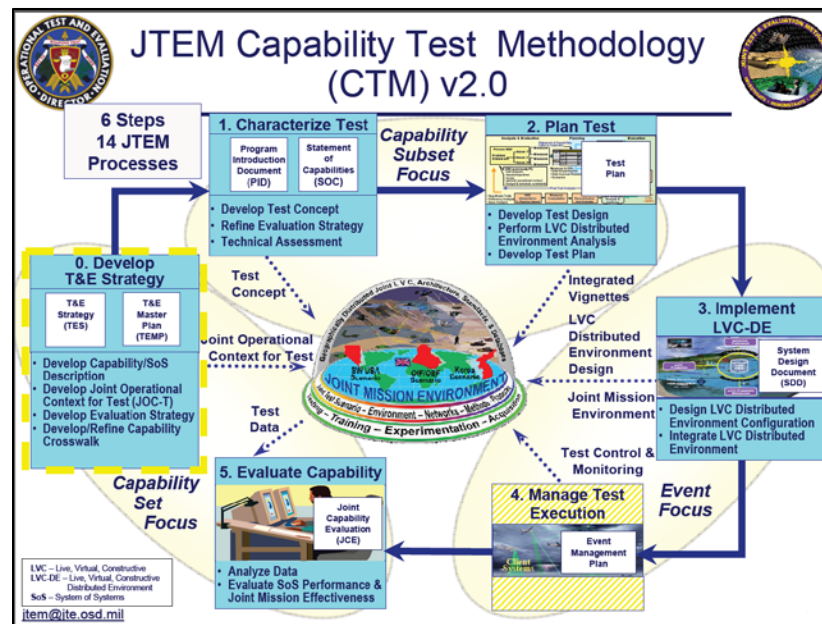


Figure 3 - JTEM Methodology (Bjorkman 2008)

A key INTEGRAL FIRE lesson learned was realizing the need for each acquisition program to identify a lead organization for LVC testing. This lead distributed test organization should coordinate directly with the program's test commander to seamlessly integrate distributed testing into the overall developmental and operational test plans. Processes for developing joint test concepts, joint operational contexts, and joint mission evaluation strategies are too important to be confined to multiple test characterizations by distributed test organizations (Bjorkman and Gray 2009b). Additional insights gained included the importance of establishing cost and responsibility

arrangements through formal relationships in order to efficiently capitalize on persistent test opportunities.

1.5 Way Ahead

Realistic joint testing on a persistent basis is a requirement for DT&E and OT&E. This is necessary to ensure problems are identified early, the fielding of viable systems are expedited to resolve capability gaps, and reasonable IOT&E expectations are established.

Achieving these goals requires the integration of distributed testing into the overall test strategy of acquisition programs. The technical infrastructure is place, the distributed federation consists of a broad and continually growing representation of joint capabilities, the methodology to organize and execute distributed test events is proven, and efforts to ensure the validity of the distributed environment are well underway.

The challenge now resides in designing distributed tests that produce statistically rigorous and actionable OT&E results. Combining geographically separated units, often conducting simultaneous tests of their own systems, to create a sufficiently robust test environment introduces new challenges. The remainder of this paper discusses the lessons learned from AGILE Fire developmental test events and highlights considerations for conducting OT&E distributed testing.

2. Case Study - AGILE Fire

2.1 AGILE Fire - Background

Daily operations throughout the CENTCOM area of responsibility require the timely and effective execution of joint fires. This often requires extensive coordination to deconflict simultaneous joint effects in a congested battlespace. Concepts to enhance the effectiveness of joint fires through new weapons systems and changes to the command and control structure are in the developmental testing process. Ongoing initiatives include the Networked Enabled Weapons (NEW) and Joint Air Ground Integration Cell (JAGIC) programs designed to provide new technical capabilities and command and control arrangements that improve the execution of joint fires.

The Air Ground Integrated Layer Explorations Fire (AGILE Fire) is the premier example of an operationally realistic virtual battlespace specifically designed to support developmental testing. AGILE Fire was established to identify joint fires interoperability gaps, shortfalls, and redundancies in current systems and network deficiencies between USAF and USA air/ground communication layers (March 2010).

The overarching objectives of the multi-phase AGILE Fire endeavor are (March 2010):

1. Provide credible analytical results for decision making.
2. Capitalize on M&S investments for maturing existing data links, and emerging command and control capabilities.
3. Focus on the interoperability within and between air and ground communication layers.

4. Capture the requirements for emerging technologies and interfaces to existing force structure in mission contexts.

The AGILE Fire virtual battlespace is constructed with entities representing those capabilities required to execute joint fires. Creating this realistic environment required the integration of units located throughout the country. AGILE Fire capitalized on the JMETC network infrastructure to integrate each facility's unique tactical or technical capability into the test scenario (see Figure-4).

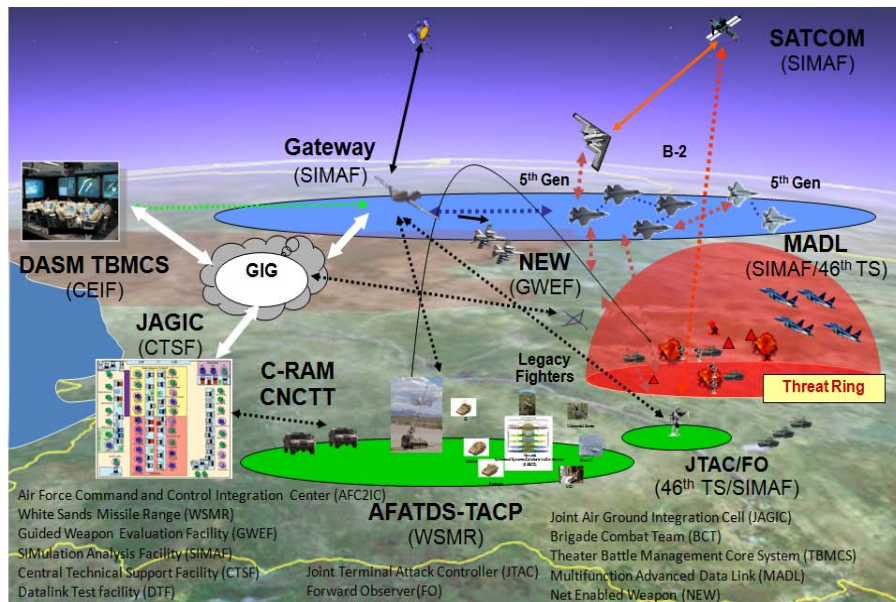


Figure 4 - AGILE Fire Battlespace (Feldman 2010)

AGILE Fire applied aspects of the JTEM Capability Test Methodology by delegating exercise design and execution responsibilities to core teams jointly lead by SIMAF and the 46th Test Squadron. These working groups included Operations Analysis, Infrastructure, Security & LVC Integrated Product Teams. Support for participating organizations included development of Measures of Effectiveness (MOE)

and Measures of Performance (MOP), design of joint mission scenarios tailored to assess these objectives, and technical expertise in LVC network connectivity and data collection (March 2010).

The AGILE Fire battlespace consisted of LVC elements from across the country (see Figure-5). The first two AGILE Fire events involved constructive and virtual participants leading to the successful integration of live JTACs tracking a live vehicle in AGILE Fire Phase-III mission threads.

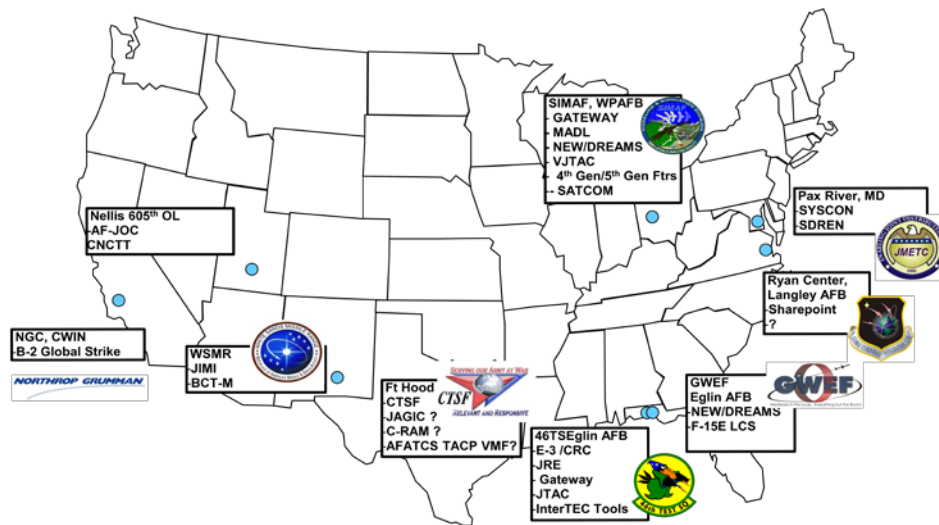


Figure 5 - AGILE Fire Phase-II Battlespace (March 2010)

The first three AGILE Fire events occurred over the period beginning January 2010 through January 2011 and included, among others, the following joint participants:

USAF

- Simulation and Analysis Facility (SIMAF)
- 46th Test Squadron / Command and Control Test Facility (Eglin AFB)
- Guided Weapons Evaluation Facility (Eglin AFB)
- 653rd Electronic Systems Wing (Hanscom AFB)

- Air Force Command and Control Integration Center (Langley AFB)

USA

- Central Test Support Facility (Ft Hood)
- White Sands Missile Range (WSMR)

2.2 *AGILE Fire Phase-I: January 2010*

The first phase of AGILE Fire centered around four test initiatives tactically relevant to the execution of joint fires (see Table-1). A closer look into two of the initiatives, NEW and JAGIC, will highlight the benefits of LVC developmental testing for new concepts and weapons systems. Tracking these two projects through the first three phases of AGILE Fire demonstrates how the ability to iteratively test in the virtual battlespace provides program managers a viable test venue to verify technical capabilities, identify program strengths and weaknesses, and implement appropriate corrective actions.

Initiative	Problem	Proposed Solution
Gateway	The inability of 5 th Generation fighters to communicate with the larger force on a similar network and the inability to pass high bandwidth data in satellite denied area to distributed surface forces.	Data Link to communicate messages from 5 th Generation to 4 th Generation aircraft. Multi-gateway implementation will allow high-bandwidth data to be passed between geographically separated surface forces.
AANI	The inability of 5 th Generation fighters to interoperate during activity in the denied access area.	Incorporate a common Data Link solution across all platforms expected to operate in the denied access environment .
NEW	Current weapons for air-to-ground (A-G) employment are unable to successfully engage targets in dynamic situations with clutter or weather impacting target track and/or ID.	Incorporate a Data Link solution across weapons and systems/nodes expected to interoperate in the dynamic targeting or Close Air Support. Current implementations allow for either Link16 control or line-of-sight (LOS) radio-based control.
JAGIC	No single C2 authority facilitating integration of air-ground operations at the lowest tactical level. Ad hoc organizations and processes. Combat effectiveness restricted and operational risk increased.	Develop modular and scalable cell to integrate and coordinate airspace. Emphasis on collocating TACS personnel with the ground element.
AANI: Advanced Aircraft Network Integration NEW: Networked Enabled Weapon JAGIC: Joint Air Ground Integration Cell		

Table 1 - AGILE Fire Phase-I Initiatives (March 2009)

Executing joint fires in a dynamic theater of operations can be a complex endeavor in time critical environments. A troops in contact scenario that requires immediate close fires support is a realistic example. Obtaining the desired joint fires effects depends on swift joint command and control coordination to clear the airspace, identifying the appropriate asset with suitable munitions and, if tactically appropriate, ensuring contact is established between the supporting asset and the Joint Terminal Attack Controller. The depiction in Figure-6 characterizes the potential complexities associated with airspace deconfliction.

Figure 6 - Airspace Structure (JAGIC 2009)

to better integrate all activities over and within the ground commanders airspace. Figure-7 depicts one potential JAGIC configuration.

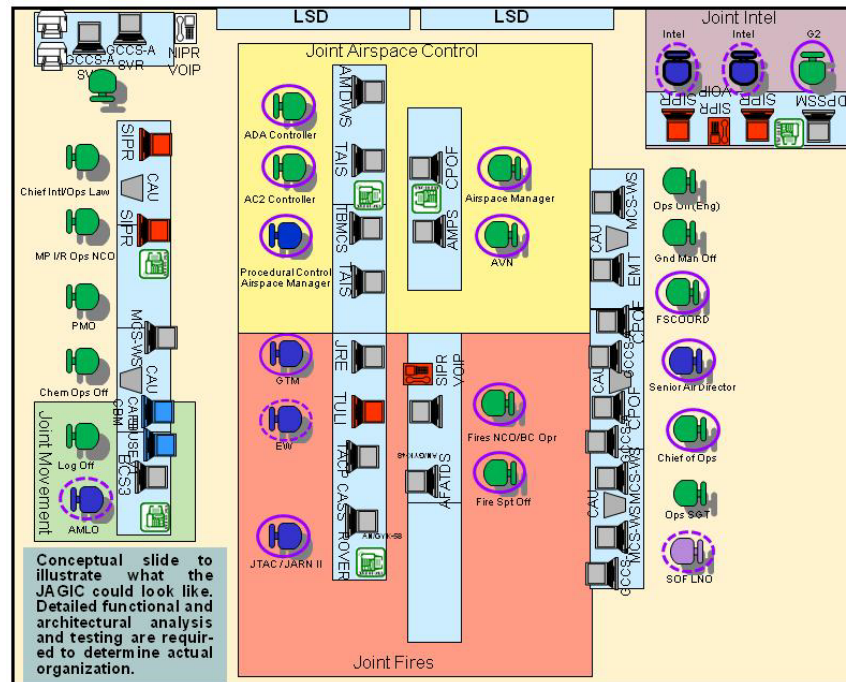


Figure 7 - Potential JAGIC Configuration (JAGIC 2009)

In 2008 both the senior level USAF CORONA conference and Army-Air Force Board General Officer Steering Committee approved staffing of the JAGIC tactical operating concept for CSAF and CSA approval.

The importance of this initiative is best described as:

Lessons learned from US combat operations repeatedly highlight significant difficulties integrating airspace control and fires deconfliction over and within a ground commander's Area of Operation (AO), particularly in areas of high density operations.

This problem is due to the significant increase in Unmanned Aircraft Systems (UAS), multiple supported commanders within the same AO,

doctrinal disconnects, the lack of reliable communications and a common operating picture resulting in ad hoc organizations and processes. Currently there is no single C2 authority/system facilitating horizontal component integration of all air-ground operations at the lowest tactical levels. The inability to integrate all airspace users, fires, air defense and air traffic control in near-real time restricts combat effectiveness, efficiency and increases risk (JAGIC 2009).

The NEW concept is designed to address the operational challenge of engaging targets in dynamic situations where environmental clutter or weather negatively impacts target tracking and/or identification (Watson 2010). A proposed solution is the real-time exchange of target information between a Joint Terminal Attack Controller and either an artillery-delivered projectile or fighter delivered air-to-ground munitions. During inclement weather conditions or high clutter environments the NEW weapon is designed to receive target updates directly from the JTAC until the final phase of the engagement.

AGILE Fire Phase-I successfully created a realistic joint operational testing environment with a level of fidelity that enabled program managers to obtain accurate system assessments. JAGIC test results verified that the proposed command and control structure is capable of meeting the program's stated goals. Additional testing was recommended in order to assess the JAGIC concept in a scenario involving a more realistic JAGIC composition of intelligence personnel and key division staff members (Allison 2010). The results for the NEW concept were also very positive and the program was considered viable for further development.

2.2 *AGILE Fire Phase-II: August 2010*

The second phase of AGILE Fire exemplifies the benefits of persistent testing in the LVC domain as envisioned in the U.S. Department of Defense (DoD) 2006-2011 Strategic Planning Guidance for Joint Testing. The ability to conduct consistent testing in a robust joint operational environment provided AGILE Fire customers the opportunity to adjust, and in some instances expand, the scope of their test objectives. A second noteworthy trend for Phase-II was an escalation in participation from four customer projects to a total of ten initiatives; clearly an ambitious load for a one-week virtual test event.

The successful USA/USAF interoperability Phase-I results led JAGIC planners to shift their objectives to the development of Tactics, Techniques and Procedures (TTP) using existing and potential near term capabilities. The original mission threads were considered suitable but the data collection plan was adjusted to capture baseline airspace management process timelines associated with dynamic Air Tasking Order (ATO) and Airspace Control Order (ACO) changes (Allison 2010). An illustration of the primary mission thread coordination elements is depicted in Figure-8.

The implications of the ability to conduct consistent robust testing to assess JAGIC test objectives should not be overlooked. These performance objectives evolved from high-level proof of concept in AGILE Fire Phase-I to more detailed and operationally applicable metrics in AGILE Fire Phase-II (Allison 2010).

JAGIC Phase-II Objectives

- Document current AF and Army JAGIC collaborations capabilities.
- Use currently fielded systems to include effectively conduct JAGIC

tasks.

- Effectively execute joint defensive counter-air missions against enemy air assets.
- Utilize an ACO with Airspace Control Measures to dynamically utilize airspace.
- Effectively integrate Joint Fires with on-going joint flying operations.

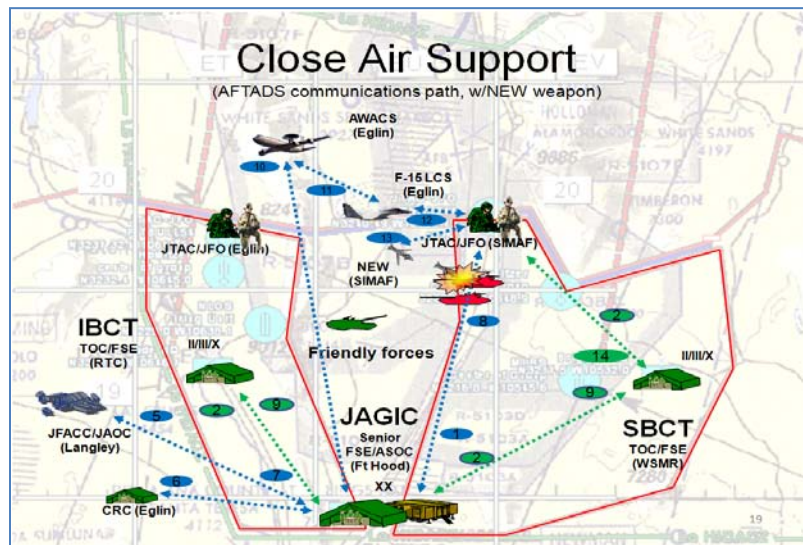


Figure 8 - Agile Fire Phase-II Close Air Support (Allison 2010)

Test objectives for Network Enabled Weapon (NEW) also evolved to capture operationally relevant metrics for both the air-to-ground and near line-of-sight (NLOS) delivery options. This included Joint Terminal Attack Controller (JTAC) control of a virtual weapon, such as a small diameter bomb (SDB), via a distributed network. Figure-9 outlines the sequence of events associated with NEW/SDB employment. Other NEW Phase-II test objectives included (March 2010):

NEW Phase-II Objectives:

- NEW CONOPS development and validation.
- Exercise and mature NEW TTPs.
- Validate proposed changes to message structures (e.g., ATO/ACO/etc).

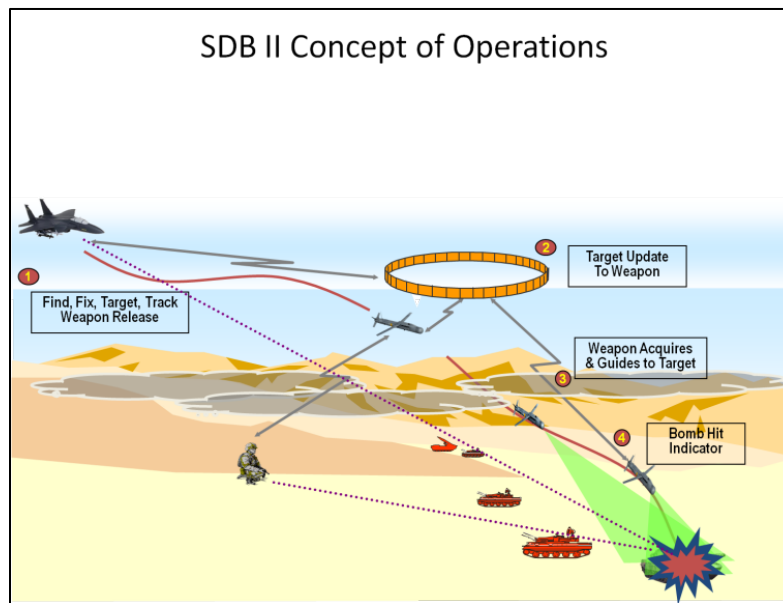


Figure 9 - NEW Small Diameter Bomb Employment (Watson 2010)

2.3 *AGILE Fire PHASE III - January 2011*

The ability to conduct persistent testing in the virtual arena again provided program managers the opportunity to analyze results, establish lessons learned and further improve system development by adjusting objectives to focus on key performance areas.

Analysis of the Phase II JAGIC results provided insights and recommendations to enhance JAGIC integration and coordination processes. To better stress these processes, the desired Phase-III mission threads included CAS, surface-to-surface fires, counter helicopter and counter unmanned aircraft missions, Combat Search and Rescue and

dynamic Airspace Control Measure (ACM) modifications. This broader range of operational scenarios provided an opportunity to validate baseline JAGIC TTPs over a broad spectrum of mission requirements (JAGIC 2010).

The Phase-III NEW objectives also built upon previous successes. Objectives centered on expanding the use of various mission planning message formats, with an emphasis on those pertaining to SDB integration concerns, and further develop the NEW Concept of Operations and TTPs with more extensive JTAC mission threads. This included operational scenarios of notable technical achievement involving a live JTAC controlling a virtual NEW against a live moving vehicle (see Figure-10).

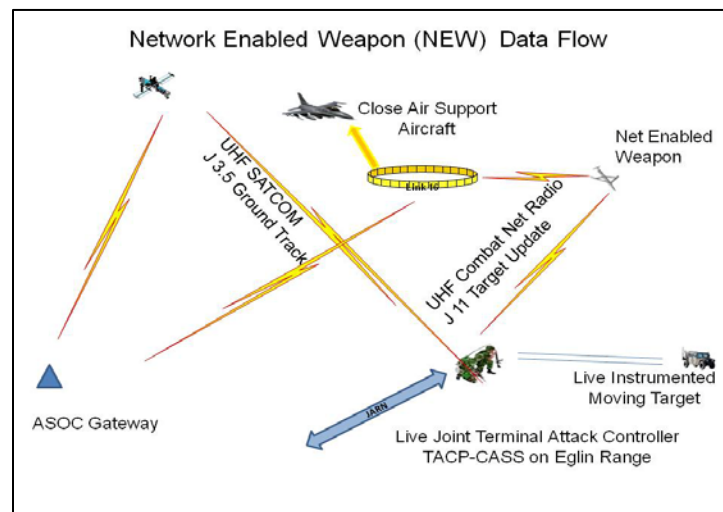


Figure 10 - NEW Data Flow (Watson 2010)

AGILE Fire Phase-III also provided the first opportunity to test the Net Enabled Weapons Controller Interface Module Situational Awareness, Analysis, and Archiving (NEWSIM SA) kit. With initial development complete AGILE Fire was the only viable venue to conduct testing involving a realistic and robust Link-16 environment involving the desired joint participants (Erickson 2010). This is an example of using joint

operational LVC testing at a point in the development process where system problems can be identified and, if necessary, the requirements refined before it becomes technically infeasible or cost prohibitive.

2.4 AGILE Fire Summary

Tracking the progression of JAGIC and NEW development through AGILE Fire events highlights the value of a joint operational LVC testing capability. Both programs were able to benefit from the opportunity to perform in a robust virtual joint environment. The ability to persistently test over a relatively short period provided program managers the opportunity to expand their objectives from the realm of technical viability to specific operational mission tasks.

3. Analysis

3.1 *Big Picture - Positive Trends*

The path to persistent and robust joint operational LVC testing is on an improving trajectory. The infrastructure is in place, the number of joint participants is growing and, possibly most importantly, there is a growing pool of personnel increasingly skilled at designing and executing LVC test events. Successful venues like AGILE Fire, Joint Expeditionary Force Experiment (JEFX) and INTEGRAL FIRE have established a foundation upon which to implement DOT&E's direction, as promulgated through the *Testing in a Joint Environment Roadmap*, to utilize the LVC joint test environment for evaluating system performance and joint mission effectiveness. These trends suggest that the test community is on the path towards establishing a persistent and robust joint LVC distributed test capability. This is a significant pre-cursor towards achieving the DoD strategic goal of "testing how we fight" (DoD 2004a).

3.2 *LVC in OT&E*

Developmental testing in the LVC environment appears fairly well defined. LVC events that contain carefully constructed mission threads and the appropriate diversity of assets provide a vehicle for refining requirements and discovering technical flaws early. The use of LVC for operational testing is not as well defined, although its utility appears well touted. Although LVC use in DT&E and experimental events is full of successes its utilization for OT&E remains largely unexplored. This study examines DT&E use of LVC and conjectures how best practices from these events, coupled with the accepted

tenets of industrial experimental design, can potentially enhance the statistical rigor of OT&E LVC events.

The following analysis takes a look at NEW test execution during AGILE Fire Phase-III. NEW testing in this event centered on program objectives designed to assess data flow and network integration in a realistic joint operational environment. Experiment structure and data collection were designed and executed with notable success using DT&E test methodology. The question of interest pertains to design of experiment considerations associated with transitioning from DT&E to OT&E LVC test events.

3.3 Conflicting Forces: Robust Joint Environment vs. Manageable Design

Conducting relevant testing in a robust joint environment involves the participation of geographically dispersed participants. As a new weapons system proceeds through the advanced phases of testing the required LVC mission threads are likely to increase in complexity. This will inherently require an escalation in the number of participants necessary to achieve the desired level of operational fidelity. As the LVC test venue matures it is also inevitable that more program managers will want to conduct testing in available events. The number of initiatives utilizing the AGILE Fire venue more than doubled in just one event iteration. This desire to capitalize on LVC test opportunities will lead to future OT&E LVC events with simultaneous testing of multiple weapon systems. This has not always been done in a statistically rigorous manner.

From a design of experiment perspective it may be difficult to capture actionable data in LVC events that involve an extensive number of parties, many of whom are

conducting their own system testing. In addition to Verification, Validation and Accreditation (VV&A) challenges to establish the LVC test environment, test designers must construct the experiment in a manner that captures relevant metrics and while accounting for the potential noise and variance introduced by other participating nodes.

Ultimately there are two potentially conflicting test requirements: the mandate to conduct testing in a "system of systems" atmosphere and the need to design the OT&E experiment so that the results provide engineers with statistically rigorous and actionable data for acquisition decisions. It is very likely that the weapons system being tested relies on the timely and accurate simultaneous participation of external systems of interest. This can create obvious problems in collecting actionable data. AGILE Fire NEW test designers have found effective ways to overcome many of these issues in DT&E test events. However, OT&E testing introduces new challenges that will require additional innovations and techniques.

3.4 Developing Experimental Measures

The design of MOEs and MOPs are so important to effective testing that AGILE Fire exercise planners publish guidance outlining techniques to develop quantifiable measures (Evans 2010). AGILE Fire Phase-III NEW overall analysis objectives were based on connectivity, data exchange with proposed message format changes to existing mission planning tools, and data flow between NEW, Command and Control, Tactical Air Control Parties and F-15E aircraft (see Table-2). A closer look at these MOE/MOPs provides insight into challenges in transitioning from DT&E to OT&E LVC test events.

The first analysis objective is constructed to capture the integration of new message formats. Daily tasking products such as the Air Tasking Order (ATO), Airspace Control Order (ACO) and OPTASKLINK adhere to very precise message structures. Existing message formats do not provide the necessary fidelity or scope of information required to execute NEW missions. Consequently, the NEW Integrated Working Group (NEWIWG) has proposed modifications to existing message formats to better support NEW employment (Watson 2010). The subsequent MOEs that assess message format integration are clearly defined with objective and measurable criteria. Each supporting MOP can also be precisely measured from exercise results.

Part 1 - Analytic Requirements					Part 2 - Collection Requirements & Analysis Methodology			
AO #	Analysis Objective	MOE/ MOP #	Measure of Effectiveness (MOE) / Measure of Performance (MOP)	Data Element	Data Source	Data Collection Method	Analysis Methodology	Data Product
1	Demonstrate NEW ATO, ACO, OPTASKLINK, OPTASK Combat Net Radio formats to execute NEW tasking	MOE 1.1	How well does the format of the proposed OPTASK CNR Support the CAS Mission Thread?				Review OPTASK CNR prior to test and after test to verify all required info included.	Report
		MOP 1.1.1	Percent of required parameters included in NEWIWG OPTASK CNR	JU for Weapon, Aircraft, JTAC	OPTASKLINK	N/A		
		MOE 1.2	How well does the NEWIWG proposed ATO format support the CAS mission thread?	Weapon JU blocks			Review ATO prior to test and after test to verify all required info included.	Report
		MOP 1.2.1	Percent of required parameters included in the NEWIWG ATO	Target Track blocks	ATO	N/A		
				Controller JU blocks	ATO	N/A		
		MOE 1.3	Integrate NEW Tasking / Planning Data Flow elements					
		MOP 1.3.1	Capture time required to develop NEW tasking Supplement	Maptool Kit or Hand Jammed	ATO	Timing		
		MOP 1.3.2	Capture time required to enter NEWINFO in AMPN Fields	Maptool Kit or Hand Jammed	ATO	Timing		
		MOP 1.3.3	Capture time associated with Taskview / JMPS Target Mgr	Maptool Kit or Hand Jammed	ATO	Timing		
2	Demonstrate the TACP-CASS and Army Fire Support Integration	MOE 2.1	How interoperable are the F-15E and Army C2 systems such as AFATDS, FOS, and PFED with respect to CAS missions?					Report
		MOP 2.1.1	Percent of messages altered during system message exchanges	J3.5	Message Traffic	Recording	Review message content for changes between TACP CASS and Weapon	
		MOE 2.2	How do test artifacts affect the thread?					Report
		MOP 2.2.1	How accurate are the IFTUs?	Distance between IFTU Coordinate and actual target location	Message Traffic	Recording	Measure distance between IFTU Coordinate and actual target location	
		MOP 2.2.2	How much latency is between the TACP CASS and DREAMS Weapon?	Time Duration for IFTU messages	Message Traffic	Recording	Measure time between IFTU Sent and IFTU received.	
3	Evaluate Airspace De-confliction (SDB II Exclusion Zones) Processes for standoff NEW's	MOE 3.1	How well do the processes for NEW support CAS missions?					Report
		MOP 3.1.1	Do any weapons violate the airspace controls?	Weapon flight track	Message Traffic	Recording	Review Airspace corridors and verify weapon does not violate zones.	
		MOP 3.1.2	How long does the airspace deconfliction process take?	Time Duration	Message Traffic	Recording	Measure time between CAS Thread step x to step y.	
		MOP 3.1.3	How long from Call for Fires to Weapon Launch?	Time Duration	Message Traffic	Recording	Measure time between CAS Thread step x to step y.	

Table 2 - NEW Analysis Traceability Matrix (McRae and Watson 2010)

The wording of MOP 1.3.1 to input data by "Maptool or Hand Jammed" does raise an important topic and potential source of variance for future OT&E LVC test events. Although not relevant in AGILE Fire due to the nature of the DT&E objective, MOPs for OT&E events may need to incorporate the inherent variance associated with human performance. To obtain accurate system employment characteristics the person involved should perform at a level consistent with that expected in an operational environment. One consideration is the level of training. A contractor intimately familiar with all aspects of the system is likely to execute duties at a higher capability than a less experienced user in an operational environment. Additionally, the performance should replicate expected operational workloads. The demands of multi-tasking in a stressful environment are a reality of combat operations. Conducting an OT&E test where the user is performing a single task in a sanitized environment is inherently going to produce results different from what will occur in a combat situation. DoD DOT&E guidance to ensure the developer and the operational community share a common understanding of the Concept of Operations (CONOPS) is particularly relevant in this regard (Gilmore 2009).

The second analysis objective for NEW Phase-III testing is to assess the integration of TACP-CASS (Tactical Air Control Party - Close Air Support System) and Army Fire Support processes. These performance measures are also precisely defined and measurable for NEW DT&E testing. However, in a future OT&E "system of systems" scenario establishing measurable criteria may be difficult. Latency among systems in the LVC environment, if the experiment is not carefully designed, can significantly skew test results. The NEW Phase-III test designers carefully placed time-

stamped messages and markers throughout the experiment. This enabled latency effects to be mathematically purged from system performance results. This proved to be viable technique for an AGILE Fire DT&E event. This approach may prove to be particularly valuable during comprehensive OT&E "system of systems" testing where the effects of latency are compounded by the performance characteristics of other dependent systems.

The final AGILE Fire Phase-III analysis objective pertained to airspace de-confliction. This objective ended up being difficult to assess within the construct and limitations of the Phase-III mission threads and will be a focal point for AGILE Fire Phase-IV. There are two core airspace issues: completing a potentially complex series of command and control steps to clear all friendly assets from the airspace, and ensuring the flight profile of a NEW weapon is in accordance with airspace restrictions. Obtaining statistically rigorous OT&E results for these MOPs may prove challenging. The length of time required to clear the airspace will depend on the complexity of the mission thread. Relocating a single Unmanned Aerial Vehicle (UAV) from a specific grid will place less stress on the command and control structure than clearing a path or re-locating assets in a high-density airspace control zone. In other words, the DOE for this objective must account for mission thread. Additionally, the NEW concept applies to a variety of munitions; from short-range SDBs to long-range Joint Air-to-Surface Standoff Missiles (JASSM). OT&E LVC testing will need to account for the characteristics of these systems, and the supporting command and control structure, in different tactical scenarios involving a gamut of airspace clearance problems.

LVC OT&E Recommendations:

- a. Objectives must be precise, measurable and directly relate to the operational requirements of the system.
- b. Experimental design planning must account for variability caused by human operators.
- c. Plan to use measurement devices that provide a direct measure of test objective satisfaction.
- d. Make use of time-stamping to help alleviate latency derived response variability. The time markers can act as statistical covariates for variance reduction efforts.

3.5 Factor Definition

The Director of OT&E places significant emphasis on designing experiments to determine the effect of factors upon measureable responses (Gilmore 2010). Test plans should contain designed experiments that consider the following:

- Quantitative mission-oriented response variables for effectiveness and suitability.
- Factors that affect measures of effectiveness and suitability.
- Experiments providing a sufficient breadth of coverage across the applicable levels of the factors.
- Methods for strategically varying factors across both developmental and operational testing.

Virtual testing has proven a viable venue to identify and address system issues early and often in the DT&E process. The fundamental technical requirement of accurately exchanging NEW targeting data has improved over the course of AGILE Fire events. As systems mature and proceed through the development process establishing appropriate experimental factors, and the operationally relevant scope of those factors, is critical for effective design of OT&E LVC events.

AGILE Fire Phase-III NEW testing was successful and provided key insights on system performance and joint integration. As the NEW concept matures, a pre-cursor to statistically rigorous OT&E results is the comprehensive identification of experimental factors and associated levels. For example, establishing a factor characterizing JTAC participation would likely be required in an AGILE Fire OT&E scenario. A mission thread could involve a live JTAC on the range, a JTAC in the simulator, or possibly a constructive JTAC element with pre-determined inputs. Effectively designing the experiment that captures these various test conditions might require a JTAC factor with three different levels.

Using the summary of AGILE Fire Phase-III NEW test events, consider a hypothetical design of experiment for future NEW OT&E scenarios (see Table-3). Suppose that test designers initially identify five factors of interest. It is feasible that the factors of interest may be Call for Fires (CFF), Static/Moving target, Weapon Type, JTAC location and target type. A full factorial design, with each factor further divided into levels that span the operational spectrum of the test environment, would require 288 runs. In an AGILE Fire type of event with approximately 30 runs available the resulting experiment design would be highly fractional. This highlights the importance of

managing the scope of the design by carefully identifying and focusing the experiment design upon those factors that indeed influence system behavior.

Summary of NEW Test Events - Chronological Order													Factor				
Day	Time	CFF Type (CRAM or JTAC)	Thread (CAS, Surface Fire, Etc...)	Target (Static/ Moving)	Service (Army or AF)	Weapon (GP or Special)	BCT Area	JTAC/JFO Site	Weapon provided by	Weapon Type	Target type (tank, building, car, etc)	Target Provided By	CFF	Static-Move	Wpn	JTAC	Target
1	1+ 15	JTAC	CAS	Moving	AF	NEW	1st BCT	JTAC 1	GWEF	NEW	Vehicle	SIMAF, OS, GWEF	0	1	0	0	0
1	1+ 40	JTAC	CAS	Moving	AF	NEW	1st BCT	JTAC 1	GWEF	NEW	Vehicle	SIMAF, OS, GWEF	0	1	0	0	0
1	3+ 20	JTAC	CAS	Moving	AF	SDBII	1st BCT	JTAC 1	GWEF	NEW	Vehicle Convoy	SIMAF, OS, GWEF	0	1	1	0	0
1	4+ 40	CRAM	CAS	Static	AF	CRAM	1st BCT	JTAC 1	GWEF	NEW	Rocket Site	CRAM	1	0	3	0	3
2	1+ 40	JTAC	CAS	Static	AF	GP	1st BCT	JTAC 1	GWEF	NEW	Dug in mortar position	SIMAF, OS, GWEF	0	0	1	0	2
2	2+ 10	JTAC	CAS	Moving	AF	NEW	1st BCT	JTAC 1	GWEF	NEW	Infantry on move	SIMAF, OS, GWEF	0	1	0	0	3
2	2+ 45	JTAC	CAS	Moving	AF	Special	1st BCT	JTAC 1	GWEF	NEW	Vehicle Convoy	SIMAF, OS, GWEF	0	1	2	0	1
2	3+ 10	JTAC	CAS	Moving	AF	Special	2nd BCT	JTAC 2	GWEF	NEW	Vehicle Convoy	SIMAF, OS, GWEF	0	1	2	1	1
2	4+ 15	JTAC	CAS	Static	AF	NEW	1st BCT	JTAC 1	GWEF	NEW	Armor Convoy	SIMAF	0	0	0	0	1
2	4+ 40	CRAM	CAS	Static	AF	CRAM	1st BCT	JTAC 1	GWEF	NEW	Mortar Site	CRAM	1	0	3	0	3
3	1+ 10	CRAM	CAS	Moving	AF	CRAM	1st BCT	JTAC 1	GWEF	NEW	Mortar Team	CRAM	1	1	3	0	3
3	1+ 25	JTAC	CAS	Moving	AF	Special	2nd BCT	JTAC 2	GWEF	NEW	Convoy	SIMAF, OS, GWEF	0	1	2	1	0
3	1+ 45	JTAC	CAS	Moving	AF	Special	1st BCT	JTAC 1	GWEF	NEW	Convoy	SIMAF, OS, GWEF	0	1	2	0	0
3	3+ 35	CRAM	CAS	Moving	Army	Special	1st BCT	JTAC 1	GWEF	NEW	Artillery	CRAM	1	1	2	0	3
3	3+ 45	JTAC	CAS	Static	AF	GP	1st BCT	JTAC 1	SIMAF (6) GWEF	GBU/NEW	Mortar site	SIMAF, OS, GWEF	0	0	1	0	3
3	3+ 15	JTAC	CAS	Moving	AF	Special	2nd BCT	JTAC 2	GWEF	NEW	Armor Convoy	SIMAF, OS, GWEF	0	1	2	1	1
3	4+ 15	JTAC	CAS	Moving	AF	GP	1st BCT	JTAC 1	GWEF	NEW	Infantry	SIMAF, OS, GWEF	0	1	1	0	2
4	1+ 10	JTAC	CAS	Moving	AF	GP	1st BCT	JTAC 1	GWEF	NEW	Armor Convoy	SIMAF, OS, GWEF	0	1	1	0	1
4	1+ 30	JTAC	CAS	Moving	AF	Special	1st BCT	JTAC 1	GWEF	NEW	Convoy	SIMAF, OS, GWEF	0	1	2	0	0
4	1+ 50	JTAC	CAS	Moving	AF	Special	1st BCT	JTAC 1	GWEF	NEW	Armor Convoy	SIMAF, OS, GWEF	0	1	2	0	1
4	2+ 00	JTAC	CAS	Static	AF	GP	1st BCT	JTAC 1	GWEF	NEW	Artillery	SIMAF, OS, GWEF	0	0	1	0	3
4	3+ 10	DT TST	TST	Moving	Best Avail	Best Avail	2nd BCT	A-Fit	GWEF	NEW	Vehicle Convoy	No	2	1	0	2	1
4	3+ 40	CRAM	CAS	Moving	Army	Special	1st BCT	JAGIC	GWEF	NEW	CRAM	CRAM	1	1	2	2	3
4	4+ 10	DT TST	TST	Moving	Best Avail	Best Avail	2nd BCT	A-Fit	GWEF	NEW	Vehicle Convoy	N	2	1	0	2	1
4	4+ 40	CRAM	CAS	Moving	Army	Special	1st BCT	JTAC 1	GWEF	NEW	Mortar Site	CRAM	1	1	2	0	3
5	1+ 10	JTAC	CAS	Static	AF	GP	1st BCT	JTAC 1	GWEF	NEW	Armor	ONESAF	0	0	1	0	3
5	1+ 40	JTAC	CAS	Moving	AF	GP	1st BCT	JTAC 1	GWEF	NEW	Vehicle Convoy	SIMAF, OS, GWEF	0	1	1	0	1
5	2+ 15	JTAC	CAS	Moving	AF	Special	1st BCT	JTAC 1	GWEF	NEW	Vehicle Convoy	SIMAF, OS, GWEF	0	1	2	0	1
5	2+ 25	CRAM	CAS	Moving	AF	Special	2nd BCT	JTAC 2	GWEF	NEW	Artillery Site	CRAM	1	1	2	1	3
5	3+ 15	JTAC	CAS	Moving	AF	Special	1st BCT	JTAC 1	GWEF	NEW	Vehicle Convoy	SIMAF, OS, GWEF	0	1	2	0	1
5	3+ 35	JTAC	CAS	Moving	AF	GP	2nd BCT	JTAC 2	GWEF	NEW	Armor convoy	SIMAF, OS, GWEF	0	1	1	1	1
Factor Levels	0-JTAC 1-CRAM 2-DT TST	CAS/TST	0-Static 1-Move	0-Short (SDBII) 1. Medium (JSOW) 2. Long (JASSM) 3. - CRAM	1st/2nd	0 - JTAC1 1 - JTAC2 2 - A-Fit	GWEF	NEW	0-Vehicle Cvy 1-Armor Cvy 2-Dug-in 3-Not Dug-in	SIMAF/OS/GWEF/ CRAM							

Table 3 - Possible NEW Factors of Interest (NEW Run Matrix 2011)

For OT&E, an additional factor describing the type of mission thread may be necessary. It is unlikely that a multi-day test event will be comprised of completely different mission scenarios for each test run. This introduces the possibility that the process of executing replicated mission threads may be "learned" by test participants thereby potentially skewing test results as the event proceeds. Scenario complexity can also affect the behavior of the systems. Command and control coordination requirements

and various weapons employment procedures are often driven by the mission type and associated degree of difficulty. Adding a mission thread factor with an adequate number of levels to capture the spectrum of core mission tasks may be required if engineers are to derive operationally accurate conclusions.

LVC OT&E Recommendations:

- a. Use fewer factors of interest in fairly focused LVC test events. Unlike DT&E, OT&E must be focused on assessing how well a system meets the requirements defined for the system (possibly refined during DT&E LVC). This focused review should mean less noise factors in the experiment which for LVC equates to less complicated events.
- b. Ensure a range of factor settings sufficient to observe a discernable effect. In statistical industrial experiments factor level settings are set to cause enough response change to detect differences over the noise. These settings come from experts familiar with the system, and inferring an expected response change in OT&E using LVC, the experiment team will need experts familiar with the system of interest and the operational environment in which it will be employed. These experts, working with the experiment planners, will need to define factor levels that yield sufficient delineation among responses to detect differences over the noise in the system responses.
- c. Mission threads may need to be considered as a factor under experimental control. As Haase points out, human operators can learn how to "game" computer systems (Haase 2011). If mission threads are learned, responses can become confounded. It therefore becomes unclear if a detected improvement

in system performance can be attributed to the system factor or operator familiarity with the LVC battlespace.

3.6 Test Discipline and Data Collection

Maintaining test discipline is a challenge in any joint test environment. Virtual testing can actually raise the level of difficulty and, with participants geographically scattered, the need for extensive pre-event coordination is essential. AGILE Fire utilized the JTEM guidelines for planning and executing a distributed test event. Coordination efforts included two planning conferences with roles and responsibilities delegated among several integrated planning teams.

As LVC testing moves into the realm of OT&E it is paramount that participating parties establish precisely defined test procedures. In a robust joint distributed environment, with multiple layers of interdependent systems, there is the potential to introduce unnecessary variance or "noise" into the test. Something as simple as different personnel conducting different roles on separate runs may influence the performance of a dependent system undergoing testing. Unplanned alterations of the LVC battlespace is not conducive towards obtaining statistically rigorous test results. It is difficult to understate the challenge in controlling mission threads, especially as they grow in operational complexity.

Clearly defined and objective AGILE Fire NEW performance measures provided engineers with insightful test results that described system behavior under various operational conditions. This is particularly true for NEW overall analysis objectives one and two that pertained to the accurate exchange of data. Metrics to measure these

objectives were collected at specific points involving an exchange of information between nodes. It is not surprising that it is substantially more difficult to capture accurate data for criteria tied to actual weapons employment across the spectrum of mission tasks in a joint operational environment. An OT&E mission thread may involve collecting information over the course of complex sequence of events within the battlespace, several of which could be occurring simultaneously. Collecting actionable data will rely on test discipline to fully implement an experimental design based on well defined MOE/MOPs, identification of factors, comprehensive levels and appropriate mission threads.

LVC OT&E Recommendations:

- a. To generate statistically rigorous results, experiments require homogeneous test materials. Using LVC in a DT&E or exercise role means flexibility (to some degree) in how the LVC is allowed to evolve over the course of the experiment. These DT&E or exercise events do not necessarily require statistically defensible results. In an OT&E experiment, the results may often require statistical rigor. Unplanned changes in the LVC environment over the course of such events will increase the error and could produce a set of results that are collectively incompatible. Thus, LVC use in OT&E roles will require changes in the conduct of LVC events.
- b. Develop suites of measures associated with test objectives during test design. An LVC test event can be instrumented. Live test ranges are already quite adept at instrumentation. Use these experiences to develop methods to instrument the LVC and tie the collected data to the responses of interest. This

collected data serves to support or refute the LVC questionnaire data often used to analyze LVC events.

3.7 Possible DOE Techniques

Detailed mathematical recommendations that alleviate all of the concerns previously raised are beyond the scope of this study. However, some previous LVC design of experiment research efforts provide generalized recommendations that may be applicable to OT&E LVC testing.

The following discussion uses the general AGILE Fire Phase-III NEW factors and levels, as categorized by this study (see Table-3), as the basis for the following alternative design of experiments discussion.

3.7.1 Full Factorial Design

A full factorial design of experiment is not a feasible option since the resulting event would require too many individual runs or mission threads. In the AGILE Fire NEW situation there are five primary factors with up to four levels per factor. A completely randomized experiment would require 288 test runs for a single replication at each setting.

3.7.2 Fractional Factorial Designs

A full factorial design often requires an extensive number of runs to provide estimates for multiple effects. Higher order interactions are generally of little interest. Fractional designs use some fraction ($1/2^k$ for some k) of the full design to reduce the

experimental design size at the expense of estimating all possible effects independently. In many cases, particularly with few factors of interest, these designs work quite well. In other cases, these designs still require too many runs. Some full factorial designs can have very complicated fractional design structures (Montgomery 2009).

The successful use of fraction factorial design is based on three key ideas:

1. Sparsity of effects principle: in a system with several variables, the process is likely to be driven primarily by some of the main effects and lower-order interactions.
2. Projection Property: designs can be projected into stronger designs in the subset of significant factors.
3. Sequential Experimentation: it is possible to combine the runs of two (or more) fractional factorials to assemble sequentially a larger design to estimate the factor effects and interactions of interest (Montgomery 2009).

The opportunity to conduct persistent robust testing should help identify the key factors, and interaction of those factors, that drive system performance. This knowledge can then be used to apply the projection property and sequential experimentation.

Therefore, as the system progresses through developmental testing it may be possible to gain further insight on the driving factors and the influencing interactions. Ultimately, a well designed TEMP should provide enough insights during DOT&E to provide a reasonable starting point for OT&E experiment design. The challenge now is how to develop TEMPs to accommodate statistically based, sequentially designed, test programs.

3.7.3 Orthogonal Arrays

Orthogonal arrays have significant potential for LVC experiments as they can accommodate mixed-level factors while maintaining the economical run size necessary in most LVC experiments (Haase 2011). An orthogonal matrix, by definition, possesses columns that are linearly independent. This property yields the useful result in that the effect estimates derived from the data are also independent.

An orthogonal test plan for the notional NEW factors and levels might consist of the arrangement shown in Table-4.

Run	CFF	Static-Move	Wpn	JTAC	Target
1	0	1	1	1	1
2	1	1	1	1	1
3	2	0	1	1	0
4	0	1	0	0	1
5	1	0	3	1	0
6	2	1	3	0	1
7	0	0	3	0	0
8	1	0	0	2	0
9	2	1	2	2	1
10	0	0	2	2	0
11	1	1	2	2	1
12	2	0	0	0	0
Factor Levels	0-JTAC 1-CRAM 2-DT TST	0-Static 1-Move	0-Short 1 - Med 2 - Long 3 - CRAM	0 - JTAC1 1 - JTAC2 2 - A-Flt	0-Vehicle Cvy 1-Armor Cvy 2-Dug-in 3-Not Dug-in

Table 4 - Notional $4^2 \times 3^2 \times 2^1$ Orthogonal Array (Bolboaca and Jantschi 2007)

A twelve run experiment is the fewest number of runs possible for a $4^2 \times 3^2 \times 2^1$ orthogonal scheme (Bolboaca and Jantschi 2007). Consequently, the power of the experiment may not be suitable for actionable decision making even though we are

obtaining independent effect estimates. If feasible, the number runs should be increased to enhance the power of the experiment.

It may also be possible, especially during the initial phases of planning and testing, to identify those factors that will not significantly affect the system (Box and Tyssedal 1996). As previously discussed, the *sparsity of effects* principle assumes that the system performance is dictated by a limited number of influencing factors, and those are generally not of the higher order interactions. Therefore, the ability to independently estimate certain factors can be dropped from the design of experiment when initial analysis reveals that certain factors are inactive (Haase 2011). Testing can then focus on the strongest and most influential orthogonal projections. Early and persistent LVC testing in the development cycle can be very useful in identifying inactive factors.

3.7.4 *Split-Plot Design*

Split plot designs are a viable alternative in situations where a completely randomized test plan is not feasible. The primary benefit of split plot designs is that they generally require less runs, but at the expense of generating results with a more complicated error structure (Haase 2011).

The process is divided into whole plots and sub-plots. Factors that are difficult, or hard to change, are assigned to the whole plot with the remaining factors assigned to the sub-plot (Jones and Nachtsheim 2009). The factors contained in the sub-plot are then randomized within the whole plot. It is generally best to assign key factors of interest to the subplots (Montgomery 2009).

A drawback to the split plot design is the resulting independent error terms for the whole plot and sub-plot (Haase 2011). The degrees of freedom for the whole plot error is usually less than that of sub-plot (e.g. #factors whole plot < #factors sub-plot). Therefore, less precise estimates can be made regarding the factor effects for factors assigned to the whole plot (Jones and Nachtsheim 2009).

For the AGILE Fire NEW case, one possible split plot design might be as described in Table-5. During the course of developmental testing it may become apparent that the JTAC and Target Type are less influential factors of interest and therefore assigned to the Whole Plot. The factors of Call for Fires, Static/Moving Target and Weapon type are assumed to be substantially more influential and are therefore allocated to the sub-plot. The influencing sub-plot factors are randomized to gain further insight into their influence on system performance.

Deciphering response signals from the noise generated by an experiment of this size is important. A useful assumption is that all higher order interaction terms are negligible (i.e., sparsity of effects). Narrowing the experimental focus on primary factors and second order interaction effects is preferable. Ideally, insight regarding interaction effects will be gleaned throughout the DT&E testing process.

The split-plot experiment as established in Table-5 considers only main effects and is therefore likely to be an overoptimistic design. This design provides levels of power for sub-plot factors ranging from .715 for long range weapons to .773 for CRAM CFF. The adequacy of these levels of power and corresponding β values are likely to be very context dependent. Furthermore, incorporating influential interactions will either decrease the power or increase in the number of runs required to achieve the desired β

values. This highlights importance of previous recommendations to identify and limit the LVC DOE to the most influential factors and levels.

Run	Whole Plot		Sub-Plot		
	JTAC	Target	CFF	Static-Move	Wpn
1	2	2	2	0	2
2	2	2	1	0	0
3	2	2	0	1	0
4	2	2	2	1	3
5	2	1	1	0	3
6	2	1	0	0	2
7	2	1	0	1	1
8	2	1	2	1	0
9	2	0	1	1	2
10	2	0	0	0	3
11	2	0	2	0	1
12	2	0	0	1	3
13	1	3	1	1	1
14	1	3	2	0	2
15	1	3	1	0	0
16	1	3	0	1	0
17	1	2	2	1	3
18	1	2	1	0	3
19	1	2	0	0	2
20	1	2	0	1	1
21	0	3	2	1	0
22	0	3	1	1	2
23	0	3	0	0	3
24	0	1	2	0	1
25	0	1	0	1	3
26	0	1	1	1	1
Factor Levels	0 - JTAC1 1 - JTAC2 2 - A-Flt	0-Vehicle 1-Armor 2-Dug-in 3-Not Dug-in	0-JTAC 1-CRAM 2-DT TST	0 - Static 1 - Move	0 - Short 1 - Med 2 - Long 3 - CRAM

Table 5 - Split-Plot Experiment

3.8 Summary

Significant progress has been achieved in developing the infrastructure, federation of distributed joint participants, and the technical expertise required to execute persistent testing in the LVC arena. AGILE Fire is an example of using the distributed environment as a viable tool in achieving the DoD mandate to conduct realistic joint testing early and often in the developmental process.

Issue	OT&E LVC Recommendation
MOE/MOP Design	<ul style="list-style-type: none"> Objectives must be precise, measureable and directly relate to the operational requirements of the system. Use measurement devices that provide a direct measure of object satisfaction. Account for variability caused by human error. Make use of time stamping to help alleviate latency derived variability.
Factor Definition	<ul style="list-style-type: none"> Avoid over-complex scenarios - limit size of event to battlespace requirements. Use fewer factors of interest in focused LVC events. Ensure a range of factor level settings adequate to detect signal over noise. Incorporate experienced LVC test engineers and operational subject matter experts in identifying factors/levels. Mission threads may need to be considered as a factor under experimental control.
Test Discipline	<ul style="list-style-type: none"> Emphasize homogenous test environment. Avoid unplanned or uncoordinated fluctuations in test environment. Establish thorough test execution 'contracts' among all participants.
Data Collection	<ul style="list-style-type: none"> Migrate live test instrumentation procedures, to the maximum extent possible, in the virtual test environment. Utilize objective test results to validate or refute potential subjective test measures (e.g. questionnaires).

Table 6 - OT&E LVC Recommendations

There are fundamental differences in the roles of DT&E and OT&E. It is paramount that OT&E testing provide statistically rigorous and actionable results (see Table-6). Accomplishing this in the distributed joint environment brings several challenges. First, test events must balance the desire to create a robust joint scenario at the expense of generating an overly complex battlespace with excessive noise and variation in the test environment. There is no doubt that as a system progresses through

the OT&E process multiple distributed participants will be required. Regardless of the size, extensive pre-coordination and test discipline is vital.

The lessons learned from decades of live testing, as well as the expanding pool of LVC expertise, must be incorporated into the design of OT&E test events. This is particularly applicable in the areas of instrumentation and data collection.

Finally, involving operational subject matter expertise provides a degree of legitimacy that would otherwise be absent. LVC tests that are executed with live or virtual personnel executing their duties in a sanitized environment or application of the new system beyond the bounds of the expected CONOPS may skew OT&E results. It is important, ultimately to the warfighter executing operational missions, that decision makers are provided a realistic assessment of system performance in a combat environment.

Distributed OT&E testing is now a technical reality. If is not employed correctly the results may be of marginal value. Applying design of experiment techniques that confront issues impeding the statistical rigor of OT&E results will enable the DoD to fully capitalize on LVC investments and the expanding pool expertise capable of executing distributed test events.

Appendix A: Air University Blue Dart

Statistically Rigorous Operational Testing in the Virtual Battlespace

Combat operations require the effective integration of weapons systems from all branches of the military. Procuring new weapons systems using a stovepipe approach has proven inefficient in resolving known capability gaps on the battlefield. Consequently, the Department of Defense has recognized the need to ensure new weapons systems are exposed to robust joint operational testing early and often in the developmental process. Recent procurement failures have highlighted the importance of conducting testing in an operational environment representative of that in which the system will be employed.

Identifying problems early in the developmental cycle provides program managers the opportunity to resolve problems or adjust system requirements before it is technically infeasible or cost prohibitive. Conducting this level of consistent and realistic testing solely with live assets is simply not feasible.

Executing the appropriate fidelity of testing in the live environment is not always a viable option. Ongoing theater operations and associated airframe limitations prevent program managers from conducting large-force test exercises on a regular basis. As a result, the Department of Defense has invested significant resources into developing a distributed networked capability for testing and training. This distributed capability links facilities from geographically separated locations to conduct virtual events with live, virtual and constructive (LVC) elements.

The LVC battlespace consists of simulation resources linked together from geographically separated locations. Constructive elements are computer-generated

entities that help fulfill mission scenario requirements (e.g., computer programmed or operator 'driving' simulated F-15E at a computer console). Virtual entities consist of a participant in a simulator trained to employ that specific tactical asset (e.g., F-15E pilot in a simulator) and live entities are manned tactical assets (e.g., airborne F-15E). All participating LVC elements are linked together through the distributed network to form a single operational environment in the virtual battlespace.

The Air Ground Integrated Layer Explorations Fire (AGILE Fire) is the premier example of an operationally realistic virtual battlespace specifically designed to support developmental testing. AGILE Fire was established to identify joint fires interoperability gaps, shortfalls, and redundancies in current systems and network deficiencies between USAF and USA air-to-ground communications. AGILE Fire has proven to be a highly effective venue for conducting developmental testing in the virtual arena.

The challenge now resides in designing distributed tests that produce statistically rigorous and actionable results during operational test and evaluation events. Combining geographically separated units, often conducting simultaneous tests of their own systems, to create a sufficiently robust test environment introduces numerous challenges. These challenges include well structured test objectives, test discipline and the overall design of the experiment. A unique issue arises in achieving a realistic operational test environment without constructing a test so large that unnecessary noise is introduced.

Distributed operational testing is clearly a technical reality. However, if is not employed correctly the results may be of marginal value. Creative application of design of experiment techniques will enable the DoD to fully capitalize on LVC investments and the expanding pool expertise capable of executing distributed test events.

The views expressed in this article are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the US Government.

Bibliography

“AGILE 10-2 Net-Enabled Weapon (NEW) Assessment.” AGILE FIRE Phase I IPC. 29 September 2009.

“AGILE Fire Phase-III NEW Run Matrix.” n. pag. <https://www.tenasda.org/display/AgileFire/e.+Analysis+IPT>. 8 February 2011.

Allison, John. “JAGIC (AGILE FIRE II Status).” AGILE FIRE II FPC. HQ Air Combat Command. 29 June 2010.

Bjorkman, Eileen A. “Joint Test and Evaluation Methodology (JTEM) Overview.” National Defense Industrial Association Modeling and Simulation Committee Meeting. 23 June 2008.

Bjorkman, Eileen A. and Frank B. Gray. “Testing in a Joint Environment 2004-2008: Findings, Conclusions and Recommendations from the Joint Test and Evaluation Methodology Project.” *ITEA Journal*, 30(1): 39-44. March 2009a.

Bjorkman, Eileen A. and Frank B. Gray. “Results of Distributed Tests with Integrated Live-Virtual-Constructive Elements: The Road to Testing in a Joint Environment.” *ITEA Journal*, 30(1): 73-83. March 2009b.

Bolboaca, Solana D. and Lorentz Jantschi. “Design of Experiments: Useful Orthogonal Arrays for Number of Experiments from 4 to 16.” *Entropy*, 9(1): 198-232. December 2007.

Box, G.E.P. and John Tyssedal. “Projective properties of certain orthogonal arrays.” *Biometrika*, 1996.

Defense Science Board. “Report of the Defense Science Board Task Force on Developmental Test & Evaluation.” Washington, D.C. May 2008.

Department of Defense 2004a. “Strategic Planning Guidance (SPG) for Joint Testing in Force Transformation.” Washington, D.C. March 2004.

Department of Defense 2004b. “Testing in a Joint Environment Roadmap.” Office of the Secretary of Defense: Director of Operational Test and Evaluation. n. pag. <http://www.dote.osd.mil/reports/TestinginaJointEnvironmentPublic111204.pdf>. November 2004.

Erickson, Royce, AFC21C/MIPI. “Net Enabled Weapons Controller Interface Module Situation Awareness, Assessment, Analysis, and Archiving (NEWCIM SA).” AGILE FIRE Capability Brief. 8 September 2010.

Evans, David. "MOE/MOS/MOP Development." TENA n. pag. <https://www.tenasda.org/AgileFire/e.+Analysis+IPT>. 21 September 2010.

Feldman, Kenton, ASC/XRA-SIMAF. "Air-Ground Integrated Layer Exploration (AGILE) Fire Phase II - Distinguished Visitor Brief. " 19 August 2010.

Ferguson, Bernard and Dave Brown. "Making the Case for Distributed Testing." *ITTEA Journal of Test Evaluation*. September 2010.

Gilmore, Dr. Michael J., Office of the Director. "Guidance on the Use of Design of Experiments (DOE) in Operational Test and Evaluation." Operational Test and Evaluation Memorandum. 19 October 2010.

Gilmore, Dr. Michael J., Office of the Director. "Test and Evaluation (T&E) Initiatives." Operational Test and Evaluation Memorandum. 24 November 2009.

Haase, Case L., Captain, USAF. *A Tailored Approach to Using Experimental Design in a Live, Virtual, Constructive, Simulation Environment*. Air University, Air force Institute of Technology, Wright-Patterson AFB OH. March 2010.

"JAGIC AGILE Fire Phase-III Capability Brief." n. pag. <https://www.tenasda.org/display/AgileFire/e.+Analysis+IPT>. December 2010.

"Joint Air Ground Integration Cell: A Concept for Joint Integrated Air-Ground C2 Operations (Rev. 41.5)." AGILE Fire Phase-I Initial Planning Conference. 3 June 2009.

"Joint Mission Environment Test Capability (JMETC) Brings Readily Available Persistent Connectivity for Joint Distributed Test Events." *Joint Mission Environment Test Capability*, n. pag. <https://tenasda.org/download/attachments/6750/JMETC-Connectivity-2011-03-16.pdf>. 24 March 2011a.

"Joint Mission Environment Test Capability (JMETC) Program." Joint Mission Environment Test Capability, n. pag. <https://www.tena-sda.org/display/JMETC/General+Info>. 3 May 2011b.

Jones, Bradley and Christopher J. Nachtsheim. "Split-plot Designs: What, why and how." *Journal of Quality Technology*, 41(4): 340-361. October 2009.

March, Walter M. "Air Ground Integrated Layer Exploration (AGILE) Fire." AGILE Fire Phase-I Final Planning Conference. 1-3 December 2009.

March, Walter M. "Air Ground Integrated Layer Exploration Fire." Phase I: Executive Summary. 26 February 2010.

March, Walt, AXC/XRA. "Air Ground Integrated Layer Exploration (AGILE)." AGILE Fire Phase-II Final Planning Conference: Bedford, Massachusetts. 29 June-1 July 2010.

McRae, Craig and Randall B. Watson. "AGILE Fire NEW Analysis Traceability Matrix." TENA n. pag. <https://www.tena-sda.org/display/AgileFire/e.+Analysis+IPT>. 29 April 2010.

Montgomery, Douglas C. *Design and Analysis of Experiments*. John Wiley & Sons, New York, NY, 7th edition, 2009.

Pearce, Steve. "Net-Enabled Weapons." AGILE Fire Spiral II Final Planning Conference, 46TSSQ(ARINC). 29 June - 1 July 2010.

"State of Reliability." Operational Test and Evaluation Memorandum for Deputy of the Secretary of Defense. *Acquisition, Technology and Defense*. 30 June 2010.

USAF Global Cyberspace Integration Center. "JEFX 09-2/3 Special Areas Report: Experiment Management." June 2009.

Watson, Randall B., 46TSSQ(ARINC). "Net-Enabled Weapons (NEW)." AGILE FIRE SPIRAL-III Initial Planning Conference. 26-28 October 2010.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 074-0188	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YYYY) 26052011		2. REPORT TYPE IDE Graduate Research Project		3. DATES COVERED (From – To) Jan 2011- May 2011	
4. TITLE AND SUBTITLE Examining the Statistical Rigor of Test and Evaluation Results in the Live, Virtual and Constructive Environment				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Wilson, James G., Major, USAF				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way, Building 640 WPAFB OH 45433-8865				8. PERFORMING ORGANIZATION REPORT NUMBER AFIT/IOA/ENS/11-06	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Dr. Catherine Warner 1400 Defense Pentagon Washington, DC 20301-1400 catherine.warner@osd.mil 703-697-7247				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT <p>The Department of Defense has mandated that weapons systems undergo persistent and realistic testing in a joint operational environment. Testing for new weapons systems is to occur early and often, in an operationally realistic environment, in order to identify and correct problems before resolution options become technically infeasible and/or cost prohibitive. Executing the appropriate fidelity of testing solely in the live environment is not always a viable course of action. Advances in distributed testing capabilities combined with the establishment of the technical infrastructure are producing a continually expanding group of distributed capable participants able to play a role in robust joint operational scenarios. However, obtaining statistically rigorous results from virtual tests, especially those pertaining to Operational Test and Evaluation, remains elusive. Several considerations associated with the Design of Experiments for virtual testing are outlined, and potential methodologies explored, with the aim to ensure rigorous and actionable results are produced when using live, virtual and constructed simulations for test purposes.</p>					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Hill, Raymond R., Civ, USAF
U	U	U	U	57	19b. TELEPHONE NUMBER (Include area code) (937) 255-6565, ext 7469 (raymond.hill@afit.edu)

